Choices in the verification of S2S forecasts and their implications for Climate Services

Andrea Manrique-Suñén[1], Nube Gonzalez Reviriego[1], Verónica Torralba[1], Nicola Cortesi[1] and Francisco J. Doblas-Reyes[1,2]

(1) Barcelona Supercomputing Center

(2) Institució Catalana de Recerca i Estudis Avançats

28/10/2020

S2S Research-to-Operation and Forecast Verification Webinar

Outlook

1. Challenges in climate services

2. Challenges in verification of S2S predictions

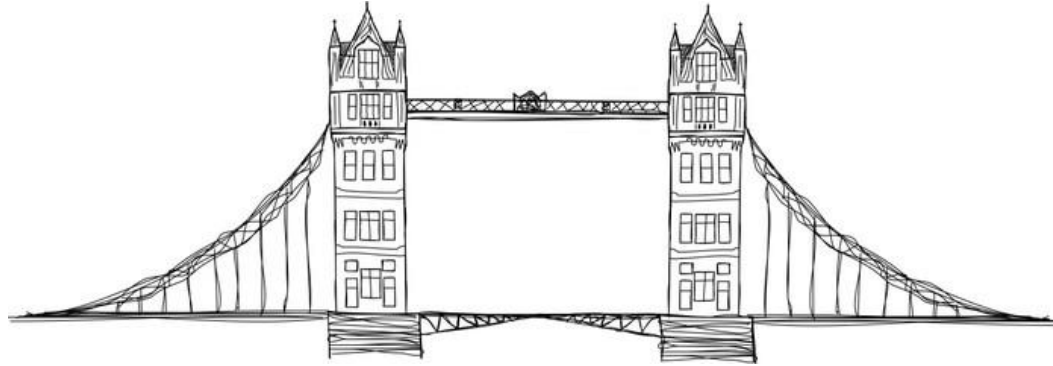3. How choices in the verification affect skill scores

Climate Services

- S2S forecast range corresponds to timing useful for many societal sectors

- The uptake of S2S predictions into actionable decisions comes with many challenges

Challenges of climate services
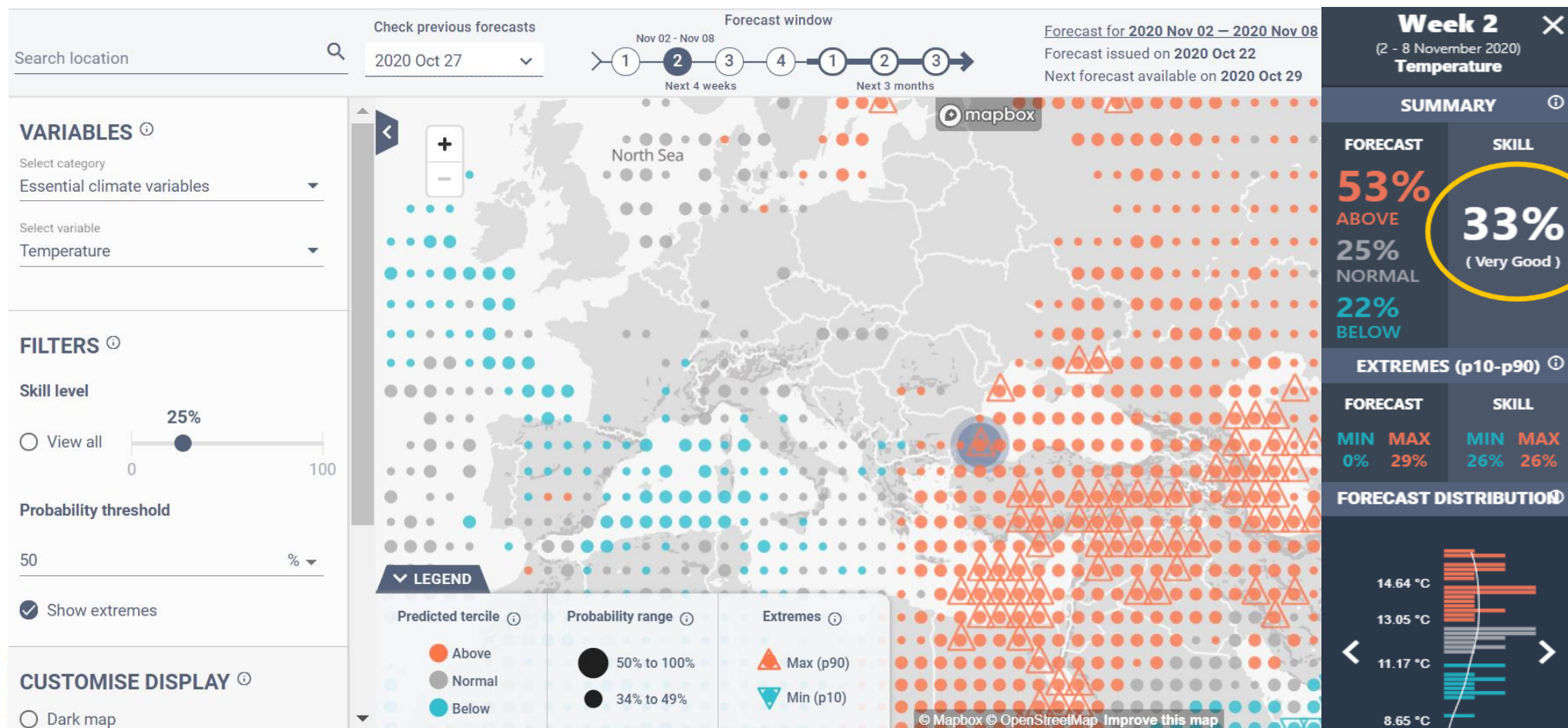
User: How much energy will I produce next month?

Scientist: Probabilistic information, skill assessment, bias adjustment, etc.

1. The **probabilistic** nature of climate predictions

   - Convey probabilistic information in a clear way (i.e. most likely tercile map)
   - Convert a probability in to an actionable decision

2. Expectations on **quality** of predictions to increase usability:

   - Provide predictions with associated information on skill
     - Skill scores, not only how good, but home much better than a reference (i.e. RPSS)
     - Seasonal and regional dependence of skill

Need for a **probabilistic verification**

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Example of a climate service:
S2S4E Project Decision Support Tool

Challenges in the verification of S2S predictions

- Heterogeneity in the subseasonal systems
- Quality measure employed
- Limited data (forecast/ hindcast)
    1. Sample size for probabilistic skill scores
    2. Definition of the climatology
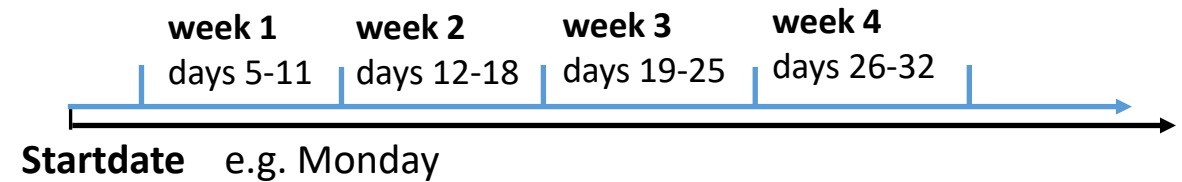   3. Application of bias adjustment

Forecast | Hindcasts

Status on 2020-10-27	Time range	Resolution	Ens. Size	Frequency	Re-forecasts	Rfc length	Rfc frequency	Rfc size
BoM (ammc)	d 0-62	T47L17	3*11	2/week	fixed	1981-2013	6/month	3*11
CMA (babj)	d 0-60	T266L56	4	2/week	on the fly	past 15 years	2/week	4
CNR-ISAC (isac)	d 0-32	0.75x0.56 L54	41	weekly	fixed	1981-2010	every 5 days	5
CNRM (lfpw)	d 0-47	T255L91	25	weekly	fixed	1993-2017	every 7 days	10
ECCC (cwao)	d 0-32	39 km L45	21	weekly	on the fly	1998-2017	weekly	4
ECMWF (ecmf)	d 0-46	Tco639/319 L91	51	2/week	on the fly	past 20 years	2/week	11
HMCR (rums)	d 0-61	1.1x1.4 L28	20	weekly	on the fly	1985-2010	weekly	10
JMA (rjtd)	d 0-33	TI479/TI319L100	50	weekly	fixed*	1981-2010	2/month	13
KMA (rksl)	d 0-60	N216L85	4	daily	on the fly	1991-2016	4/month	3
NCEP (kwbc)	d 0-44	T126L64	16	daily	fixed	1999-2010	daily	4
UKMO (egrr)	d 0-60	N216L85	4	daily	on the fly	1993-2016	4/month	7

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Subseasonal-to-Seasonal S2S Prediction Project

Verification setup in our study

- ECMWF-Ext-ENS 2016 System

- Hindcast period: 1996-2015, 11 members

- Reference: ERA- Interim reanalysis

- Bias adjustment: Simple bias correction

- Target forecast weeks

- Fair RPSS for tercile categories / Fair CRPSS

$$y_{ij} = (x_{ij} - \bar{x}) \frac{\sigma_{ref}}{\sigma_e} + \bar{o}$$

y_{ij} adjusted forecast, x_{ij}
j: member, i: year

week 1	week 2	week 3	week 4
days 5-11	days 12-18	days 19-25	days 26-32

Startdate   e.g. Monday

$$SS = \frac{S_{fcst} - S_{clim}}{S_{perf} - S_{clim}}$$

$\begin{cases} SS > 0 \text{ Forecast is better than climatology} \\ SS < 0 \text{ Forecast is worse than climatology} \end{cases}$

- Sample size for skills cores: 2 options

- Definition of climatology: 3 options

Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación

Choices in sample size for the skill score and definition of climatology
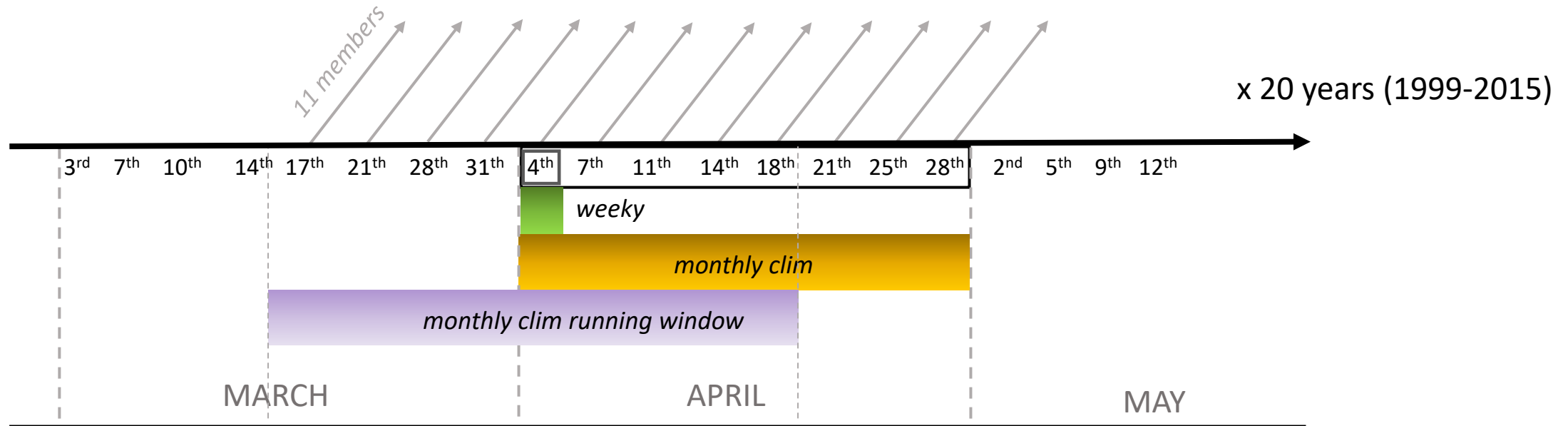
SKILL SCORE:
- Single start date: 1 start date, 20 years
- Monthly start dates: 8/9 start dates, 20 years

CLIMATOLOGY:
Weekly: 1 start date, 20 years
Monthly: All start dates in a calendar month, 8/9 start dates, 20 years
Monthly running window: Running window with 4 start dates before and after the target week, 9 start dates, 20 years

Climatology used for:
- Reference for anomalies
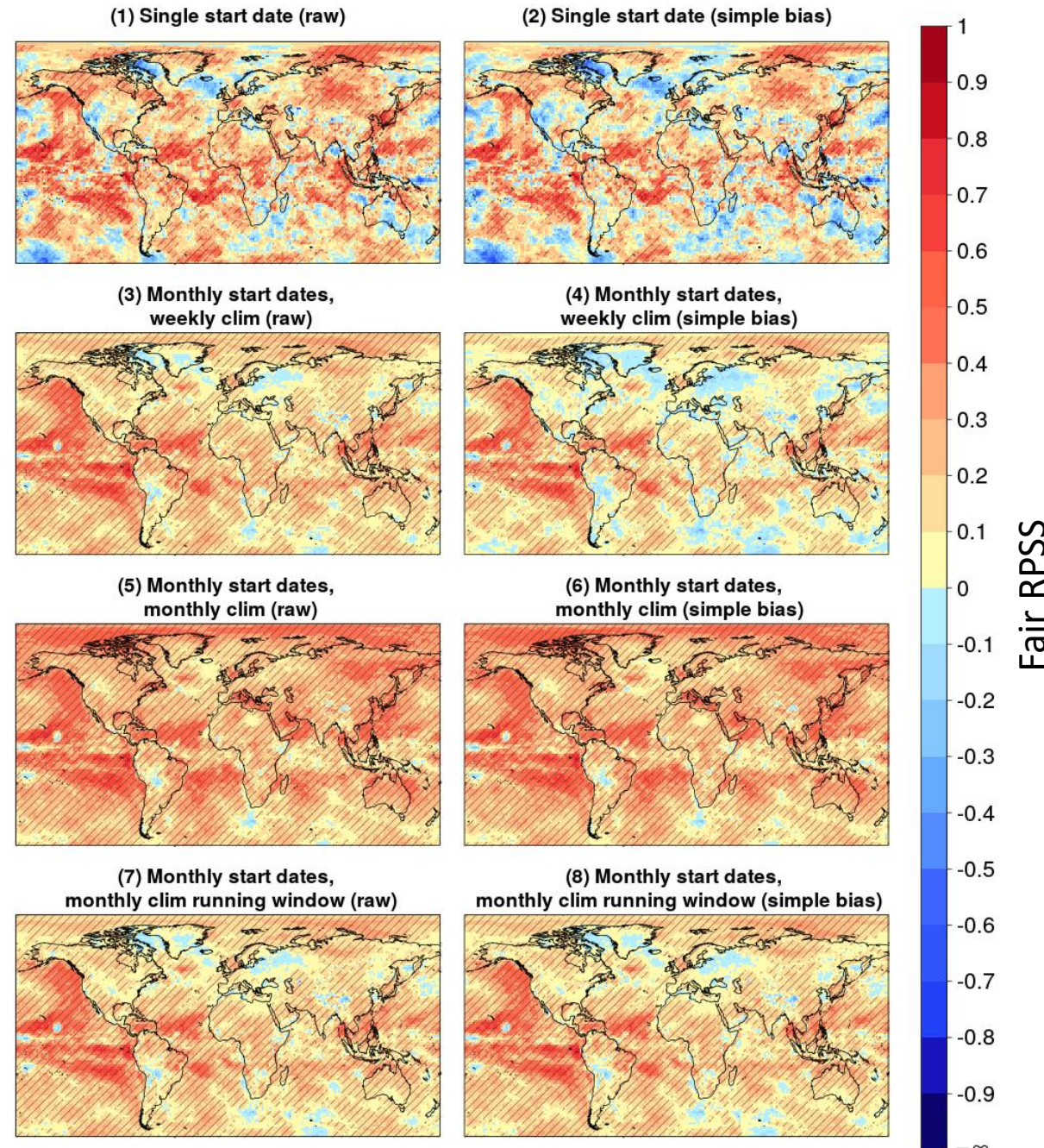- Benchmark forecast
- Bias adjustment

Verification setups

Methodology	Skill score	Climatology	Bias correction
1	**Single start date** (1 start date)	**Weekly** (1 start date)	Raw
2	**Single start date** (1 start date)	**Weekly** (1 start date)	Simple bias adjustment
3	**Monthly** (concatenating 8/9 start dates)	**Weekly** (1 start date)	Raw
4	**Monthly** (concatenating 8/9 start dates)	**Weekly** (1 start date)	Simple bias adjustment
5	**Monthly** (concatenating 8/9 start dates)	**Monthly** (8/9 start dates)	Raw
6	**Monthly** (concatenating 8/9 start dates)	**Monthly** (8/9 start dates)	Simple bias adjustment
7	**Monthly** (concatenating 8/9 start dates)	**Monthly running window** (9 start dates)	Raw
8	**Monthly** (concatenating 8/9 start dates)	**Monthly running window** (9 start dates)	Simple bias adjustment

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Same hindcast, 4 ways to perform forecast quality assessment (fair RPSS)
Var: 2m temperature

Concatenating startdates
8/9*20 yrs
data-obs pairs

April Fair RPSS terciles - Fcst time: Days 12-18

(1) Single start date (raw)

(2) Single start date (simple bias)

(3) Monthly start dates, weekly clim (raw)

(4) Monthly start dates, weekly clim (simple bias)

(5) Monthly start dates, monthly clim (raw)

(6) Monthly start dates, monthly clim (simple bias)

(7) Monthly start dates, monthly clim running window (raw)

(8) Monthly start dates, monthly clim running window (simple bias)

Fair RPSS

Single start date:
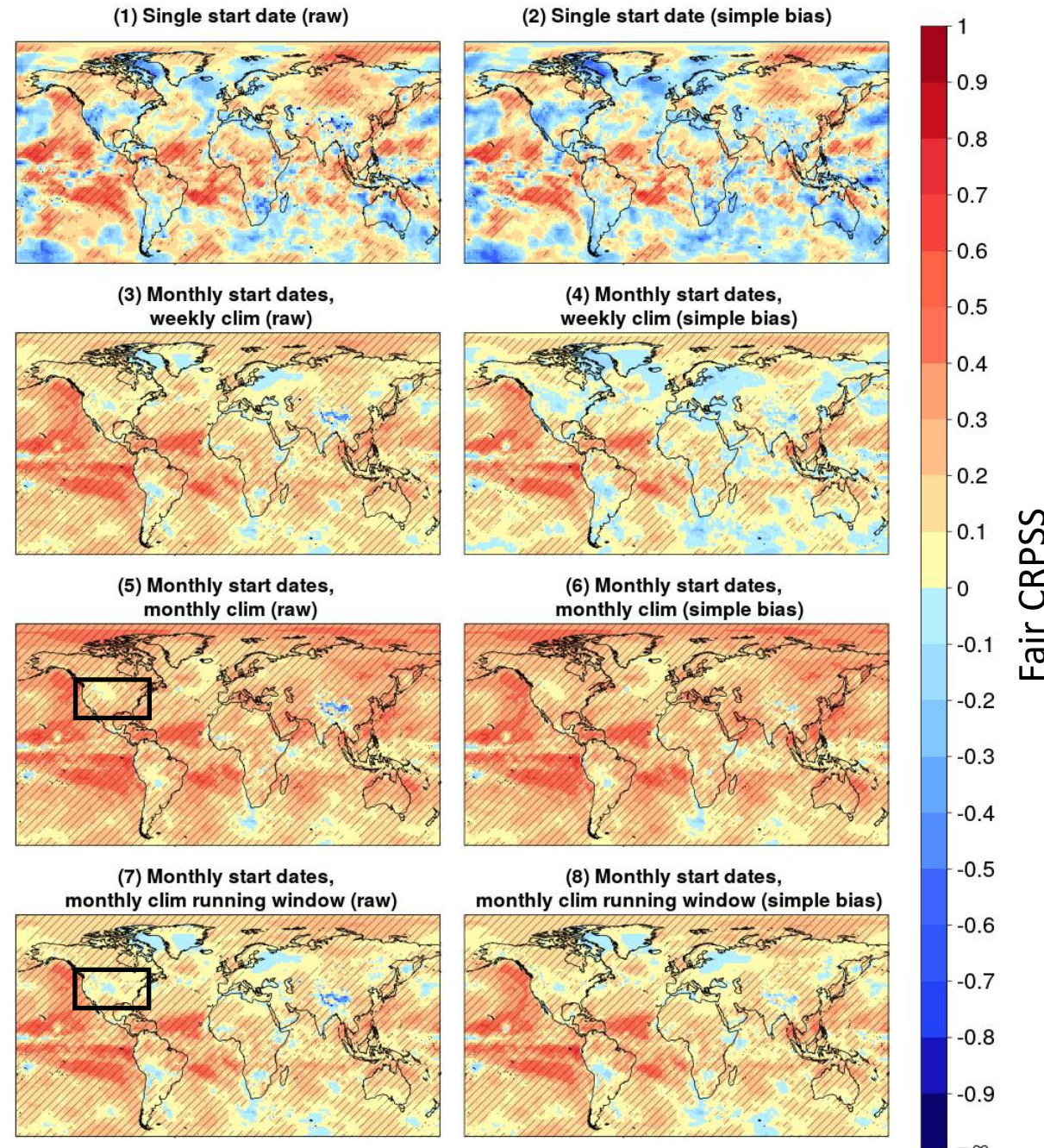Too noisy to compute skill score (20 data-obs pairs)

Weekly climatology
Simple bias adjustment degrades skill

Monthly climatology
Apparently good skill but ...
(... next slide)

Monthly climatology running window
More robust climatology for bias adjustment – less degradation

Same hindcast, 4 ways to perform forecast quality assessment (fair CRPSS)
Var: 2m temperature

Concatenating startdates
8/9*20 yrs data-obs pairs

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

April Fair CRPSS - Fcst time: Days 12-18

(1) Single start date (raw)
(2) Single start date (simple bias)
(3) Monthly start dates, weekly clim (raw)
(4) Monthly start dates, weekly clim (simple bias)
(5) Monthly start dates, monthly clim (raw)
(6) Monthly start dates, monthly clim (simple bias)
(7) Monthly start dates, monthly clim running window (raw)
(8) Monthly start dates, monthly clim running window (simple bias)

Fair CRPSS

Single start date:
Too noisy to compute skill score (20 data-obs pairs)

Weekly climatology
Simple bias adjustment degrades skill

Monthly climatology
Apparently good skill but ...
... (next slide)

Monthly climatology running window
More robust climatology for bias adjustment – less degradation

Annual evolution for a region in North America, for forecast days 12-18
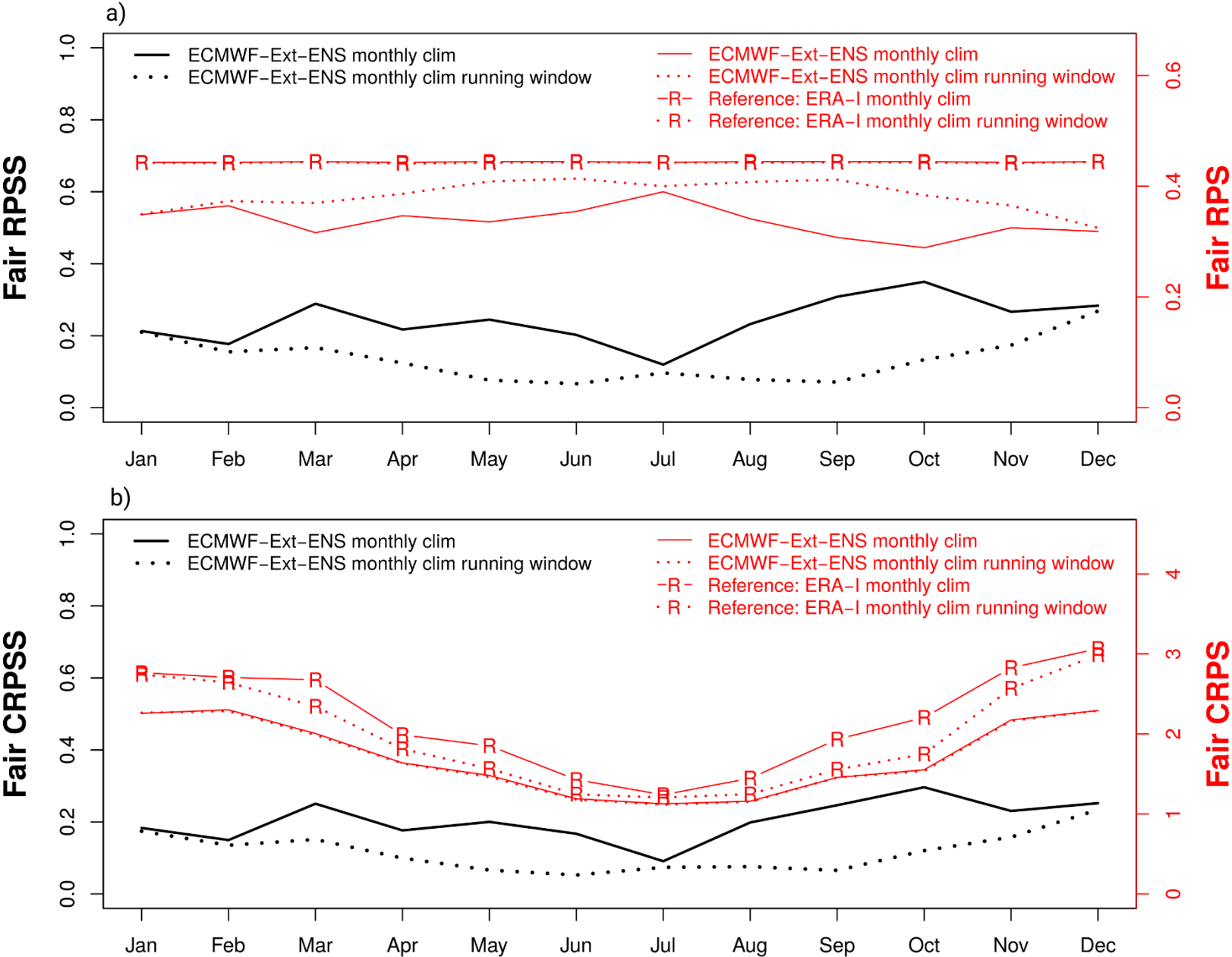
Conclusions

- SAMPLE SIZE: One single start date is not enough for a robust skill score with a 20 years hindcast. Concatenating several start dates is a good approach to increase the sample size and produce robust skill scores

- BIAS ADJUSTMENT: The reference climate distribution to bias adjust weekly averages should span a longer period than one month

- CLIMATOLOGY: The aggregation period to compute the climatology should be centered around the target week. Using a calendar month to compute climatology can lead to "artificial skill" when computing skill scores.

--> It is fundamental to carefully document all verification procedures !

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Thank you, Questions?

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Publication:

Manrique-Suñén, A., N. Gonzalez-Reviriego, V. Torralba, N. Cortesi, and F. J. Doblas-Reyes, 2020: Choices in the Verification of S2S Forecasts and Their Implications for Climate Services. *Mon. Wea. Rev.*, **148**, 3995–4008, https://doi.org/10.1175/MWR-D-20-0067.1.

S2S4E
Climate Services for Clean Energy

andrea.manrique@bsc.es